

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2006, 18:07:43 ; Search time 2527.47 Seconds
(without alignments)
703.434 Million cell updates/sec

Title: US-09-669-187A-148
Perfect score: 38
Sequence: 1 ggggacttcgctgggacttccaggggactttcc 38

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 179606

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database :
EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_est4: *
5: gb_est5: *
6: gb_est6: *
7: gb_est7: *
8: gb_est8: *
9: gb_est9: *
10: gb_est10: *
11: gb_est11: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16.2	42.6	46	1	AI544460 fb75b10.x
2	16	42.1	25	1	AA878803 of86e04.s
3	16	42.1	30	10	AG198773 Pan trogl
4	16	42.1	49	9	CC178684 RST325 Ba
5	15.6	41.1	37	8	R75772 YL21a11.r1
6	15.4	40.5	28	1	AI416657 sal0b04.x
7	15.4	40.5	34	1	AI357237 GX63a08.x
8	15.4	40.5	34	9	AZ760788 IM0554F14
9	15.4	40.5	50	10	CG800335 1118010B1
10	15	39.5	42	10	CG776618 1123002C0
11	14.6	38.4	46	1	AI917121 ts5a12.x
12	14.6	38.4	50	1	AI102736 AU102736
13	14.6	38.4	50	1	AU106243 AU106243
14	14.4	37.9	38	10	CL878011 abf23b12.
15	14.4	37.9	42	9	BZ288740 SALK_0221
16	14.4	37.9	45	9	AZ345648 IM0080K22
17	14.4	37.9	46	1	AA504387 aa59c05.r
18	14.4	37.9	50	1	AU106577 AU106577
19	14.2	37.4	35	10	CZ442924 IBBJG02.f
20	14.2	37.4	37	11	AL497861 T. brucei
21	14.2	37.4	40	1	AI039253 ox33b01.s
22	14.2	37.4	50	1	AU104208 AU104208

23	14.2	37.4	50	9	CC459460
24	14.2	37.4	50	10	AJ587887
25	14	36.8	34	1	AA931137
26	14	36.8	38	10	AG190424
27	14	36.8	40	9	AZ595836
28	14	36.8	40	9	AZ664762
29	14	36.8	40	10	CZ487957
30	14	36.8	41	9	CC200179
31	14	36.8	45	10	CL640557
32	14	36.8	46	1	AA897403
33	14	36.8	49	1	AI357381
34	14	36.8	50	2	BE311372
35	13.8	36.3	28	1	AI432930
36	13.8	36.3	33	9	AZ628058
37	13.8	36.3	37	1	AI579983
38	13.8	36.3	41	11	DR4832S
39	13.8	36.3	43	1	AI520679
40	13.8	36.3	46	1	AI223998
41	13.8	36.3	46	1	AI544460
42	13.8	36.3	46	3	BJ057570
43	13.8	36.3	50	1	AU102496
44	13.8	36.3	50	1	AU107397
45	13.8	36.3	50	11	CR178464
46	13.6	35.8	32	9	BH812215
47	13.6	35.8	40	9	AZ775335
48	13.6	35.8	41	11	TA26H11Q
49	13.6	35.8	45	8	W10989
50	13.6	35.8	46	1	AI635210
51	13.6	35.8	50	1	AU102500
52	13.6	35.8	50	1	AU102723
53	13.6	35.8	50	1	AU102932
54	13.6	35.8	50	1	AU103836
55	13.6	35.8	50	1	AU105393
56	13.6	35.8	50	1	AU106102
57	13.6	35.8	50	1	AU106131
58	13.4	35.3	26	8	H49860
59	13.4	35.3	36	9	AZ664706
60	13.4	35.3	37	1	AA871389
61	13.4	35.3	37	1	AI889881
62	13.4	35.3	42	10	CL520536
63	13.4	35.3	43	1	AI917027
64	13.4	35.3	45	1	AU254285
65	13.4	35.3	46	1	AI677817
66	13.4	35.3	46	9	BZ384414
67	13.4	35.3	46	11	TA184D05Q
68	13.4	35.3	49	1	AI597951
69	13.4	35.3	49	1	AI670105
70	13.4	35.3	50	1	AU106830
71	13.4	35.3	50	1	AU106833
72	13.2	34.7	28	1	AA721232
73	13.2	34.7	28	1	AI357300
74	13.2	34.7	33	10	CZ474868
75	13.2	34.7	34	3	BH832750
76	13.2	34.7	37	1	AA872089
77	13.2	34.7	38	1	AA903624
78	13.2	34.7	37	9	AZ773132
79	13.2	34.7	42	9	AZ331072
80	13.2	34.7	42	9	AZ983516
81	13.2	34.7	43	9	BH851950
82	13.2	34.7	46	1	AI624752
83	13.2	34.7	46	8	N54785
84	13.2	34.7	49	1	AI580812
85	13.2	34.7	49	9	AZ872239
86	13.2	34.7	49	9	BH851976
87	13.2	34.7	50	1	AU105244
88	13.2	34.7	50	9	AZ854413
89	13.2	34.7	50	9	BH814943
90	13.2	34.7	50	10	CW404207
91	13.2	34.7	50	11	CR148320
92	13.2	34.7	50	11	CR148320
93	13	34.2	27	10	CG733938
94	13	34.2	31	1	AI916689
95	13	34.2	31	2	BF213605

CC459460	SALK_1299
AJ587887	Arabidops
AA931137	oc70B07.s
AG190424	Pan trogl
AZ595836	IM0408008
AZ664762	IM0545P18
CZ487957	F05395-3p
CC200179	KG205 Bay
CL640557	G076G08 G
AA897403	al48A05.s
AI357381	gyl3B05.x
BE311372	601057775
AI432930	th44G11.x
AZ628058	IM0476122
AI579983	tg45A05.x
AL582948	Danio rer
AI520679	tl06C03.x
AI223998	qx12H07.x
AI544460	fb75B10.x
BJ057570	BJ057570
AU102496	AU102496
AU107397	AU107397
CR178464	Forward s
BH812215	SALK_0614
AZ775335	2M0007G07
AI452967	T. brucei
W10989	ma41b12.r1
AI635210	tz22S06.x
AU102500	AU102500
AU102723	AU102723
AU102932	AU102932
AU103836	AU103836
AU105393	AU105393
AU106102	AU106102
AU106131	AU106131
H49860	yo24e12.s1
AZ664706	IM0545E15
AA871389	vm34Q05.r
AI889881	Wm64F02.x
CL520536	SAK2807 F
AI917027	ts51a12.x
AU254285	AU254285
AI677817	WC80904.x
BZ384414	SALK_1355
TA184D05Q	
AI597951	ts04806.x
AI670105	wc11H03.x
AU106830	AU106830
AU106833	AU106833
AA721232	nz72B06.s
AI357300	qx15b10.x
CZ474868	d06009-5p
BH832750	603082396
AA872089	g112c10.s
AA903624	OK59F07.s
AZ773132	IM0584C06
AZ331072	IM0056A13
AZ983516	2M0264O09
BH851950	SALK_0737
AI624752	ts44801.x
N54785	YV31a05.s1
AI580812	ts11a09.x
AZ872239	2M0185C03
BH851976	SALK_0738
AU105244	AU105244
AZ854413	2M0158C01
BH814943	SALK_0672
CW404207	fb50001f0
CR148320	Forward s
CR148320	Forward s
CG733938	1119161B1
AI916689	tu89F07.x
BF213605	601844619

C 96	13	34.2	33	9	AZ424092	106203D03	169	12.6	33.2	50	1	AUI02494
C 97	13	34.2	35	1	AI590768	tw2508.x	170	12.6	33.2	50	1	AUI02495
C 98	13	34.2	36	10	AJ599818	Arabidops	171	12.6	33.2	50	1	AUI02497
C 99	13	34.2	39	8	T49679	ya7bf04.sl	172	12.6	33.2	50	1	AUI02499
C 100	13	34.2	40	8	CV933307	Pmpcm.060	173	12.6	33.2	50	1	AUI02504
C 101	13	34.2	40	10	CZ918363	4021009B0	174	12.6	33.2	50	1	AUI02508
C 102	13	34.2	43	1	AI421857	tf55e11.x	175	12.6	33.2	50	1	AUI02509
C 103	13	34.2	43	1	AI590595	tw23c11.x	176	12.6	33.2	50	1	AUI02510
C 104	13	34.2	43	1	AI613476	ty37803.x	177	12.6	33.2	50	1	AUI03928
C 105	13	34.2	46	8	W47560	zc35a01.rl	178	12.6	33.2	50	1	AUI07888
C 106	13	34.2	47	9	BH610050	KG00327-5	179	12.6	33.2	50	1	AUI07889
C 107	13	34.2	50	1	AUI05329	AUI05329	180	12.6	33.2	50	1	AUI08000
C 108	13	34.2	50	1	AUI05338	AUI05338	181	12.6	33.2	50	1	AUI08001
C 109	13	34.2	50	1	AUI05340	AUI05340	182	12.6	33.2	50	1	AUI08002
C 110	13	34.2	50	1	AUI05355	AUI05355	183	12.6	33.2	50	2	BG391762
C 111	13	34.2	50	1	AUI05358	AUI05358	184	12.6	33.2	50	9	AZ366252
C 112	13	34.2	50	1	AUI05378	AUI05378	185	12.6	33.2	50	10	AG227817
C 113	13	34.2	50	1	AUI05735	AUI05735	186	12.6	33.2	50	11	CR043223
C 114	13	34.2	50	1	AUI05735	AUI05735	187	12.4	32.6	22	1	AA931331
C 115	13	34.2	50	1	AUI05992	AUI05992	188	12.4	32.6	25	9	BZ378205
C 116	13	34.2	50	1	AUI06832	AUI06832	189	12.4	32.6	28	1	AA152566
C 117	12.8	33.7	27	9	AZ473371	IM0289A18	190	12.4	32.6	30	9	AZ788334
C 118	12.8	33.7	36	9	AZ352257	IM0090G11	191	12.4	32.6	30	10	CZ472881
C 119	12.8	33.7	36	9	AZ466529	IM0277K18	192	12.4	32.6	34	10	CZ476686
C 120	12.8	33.7	37	9	AI601822	fc12d01.x	193	12.4	32.6	37	1	AA921165
C 121	12.8	33.7	37	9	AZ346218	IM0081D16	194	12.4	32.6	37	1	AA99210
C 122	12.8	33.7	37	11	CT011645	KT011645	195	12.4	32.6	37	2	BF211603
C 123	12.8	33.7	38	3	BJ081977	BJ081977	196	12.4	32.6	38	10	AJ622374
C 124	12.8	33.7	41	7	AE683751	AJ683751	197	12.4	32.6	40	1	AI924099
C 125	12.8	33.7	42	10	CL520536	SAK2E07.F	198	12.4	32.6	42	6	W87969
C 126	12.8	33.7	46	1	AA948496	on53b06.s	199	12.4	32.6	42	8	CA797449
C 127	12.8	33.7	46	1	AI284041	qt72d04.x	200	12.4	32.6	43	9	AZ838833
C 128	12.8	33.7	46	1	AI787001	uj54e01.x	201	12.4	32.6	43	9	AZ831953
C 129	12.8	33.7	46	10	AG202640	Fan.trog1	202	12.4	32.6	46	1	AA869036
C 130	12.8	33.7	46	10	AG202640	Fan.trog1	203	12.4	32.6	46	1	AI619473
C 131	12.8	33.7	49	1	AA663894	ae74c09.s	204	12.4	32.6	46	1	AI619473
C 132	12.8	33.7	49	1	AI005180	ou60a07.x	205	12.4	32.6	48	9	AZ810742
C 133	12.8	33.7	50	1	AUI04206	AUI04206	206	12.4	32.6	48	9	AZ810742
C 134	12.8	33.7	50	1	AUI04856	AUI04856	207	12.4	32.6	49	7	CK812785
C 135	12.8	33.7	50	1	AUI05372	AUI05372	208	12.4	32.6	49	7	CK812785
C 136	12.8	33.7	50	1	AUI05730	AUI05730	209	12.4	32.6	50	1	AUI04860
C 137	12.8	33.7	50	1	AUI07888	AUI07888	210	12.4	32.6	50	1	AUI05386
C 138	12.8	33.7	50	1	AUI07889	AUI07889	211	12.4	32.6	50	1	AUI05523
C 139	12.8	33.7	50	9	AZ654289	IM0528E22	212	12.4	32.6	50	9	AZ790119
C 140	12.6	33.2	23	9	AZ585259	IM0390O14	213	12.2	32.1	31	1	AI286975
C 141	12.6	33.2	33	10	CZ469762	co5149-5p	214	12.2	32.1	31	10	CZ472358
C 142	12.6	33.2	34	1	AI811024	tr03a03.x	215	12.2	32.1	35	11	TA275003P
C 143	12.6	33.2	37	1	AI597737	tu91b01.x	216	12.2	32.1	37	1	AI357445
C 144	12.6	33.2	37	1	AI597737	tu91b01.x	217	12.2	32.1	37	1	AA275360
C 145	12.6	33.2	38	9	AZ826498	2M0102N19	218	12.2	32.1	37	9	AZ343423
C 146	12.6	33.2	40	1	AA810284	od14h07.s	219	12.2	32.1	39	10	CG724099
C 147	12.6	33.2	40	1	AI020727	ua98h09.x	220	12.2	32.1	40	1	AA428089
C 148	12.6	33.2	41	8	D38738	HUMCIET91B	221	12.2	32.1	40	9	AZ806050
C 149	12.6	33.2	42	9	CC456976	SALK.1021	222	12.2	32.1	41	9	AZ806050
C 150	12.6	33.2	43	1	AI289991	qc24d05.x	223	12.2	32.1	42	1	AJ663495
C 151	12.6	33.2	43	1	AI589982	tt72d06.x	224	12.2	32.1	42	3	BJ079239
C 152	12.6	33.2	44	3	BJ040920	BJ040920	225	12.2	32.1	42	9	AZ807828
C 153	12.6	33.2	44	6	CA794547	Cac.BL.14	226	12.2	32.1	42	10	CZ472457
C 154	12.6	33.2	44	10	AG193478	pan.trog1	227	12.2	32.1	42	10	CG776618
C 155	12.6	33.2	45	9	AZ465901	IM0276H09	228	12.2	32.1	45	3	BJ067044
C 156	12.6	33.2	46	1	AI958214	fc92f08.y	229	12.2	32.1	45	8	T67260
C 157	12.6	33.2	46	1	AJ691901	AJ691901	230	12.2	32.1	45	9	AZ345712
C 158	12.6	33.2	46	8	N40673	ym78g11.rl	231	12.2	32.1	45	9	BH851251
C 159	12.6	33.2	47	3	BI826451	603076143	232	12.2	32.1	46	1	AI554503
C 160	12.6	33.2	47	8	N62591	zal3h10.l81	233	12.2	32.1	46	9	AZ454102
C 161	12.6	33.2	47	9	BH850536	BM516312	234	12.2	32.1	46	10	CL640369
C 162	12.6	33.2	48	10	BU126913	BM516312	235	12.2	32.1	47	5	BQ593365
C 163	12.6	33.2	49	1	AU253477	AU253477	236	12.2	32.1	48	2	BG685213
C 164	12.6	33.2	49	3	BM516312	kj68e06.y	237	12.2	32.1	49	1	AA792088
C 165	12.6	33.2	49	3	BM566930	kj07h07.y	238	12.2	32.1	49	1	AA442185
C 166	12.6	33.2	50	1	AUI02486	AUI02486	239	12.2	32.1	50	1	AL792018
C 167	12.6	33.2	50	1	AUI02489	AUI02489	240	12.2	32.1	50	1	AUI04323
C 168	12.6	33.2	50	1	AUI02492	AUI02492	241	12.2	32.1	50	1	AUI05719

242	12.2	32.1	50	1	AUI06350	AUI06350
243	12.2	32.1	50	1	AUI06889	AUI06889
244	12.2	32.1	50	1	AUI06925	AUI06925
245	12.2	32.1	50	6	CA969709	CCLX0624
246	12.2	32.1	50	9	AZ918144	100600300
247	12.2	32.1	50	10	CZ470013	C05470D-3
248	12	31.6	22	1	AA985475	OQ96E09.s
249	12	31.6	25	1	AW246315	2821832.5
250	12	31.6	26	9	AZ333170	1M0062B07
251	12	31.6	26	9	AZ615102	1M0444G15
252	12	31.6	32	10	CZ194681	PSI7048-N
253	12	31.6	33	10	CG731153	1119134F1
254	12	31.6	34	9	BH789672	SALK 0423
255	12	31.6	34	9	BZ764154	SALK 1240
256	12	31.6	34	9	BZ764155	SALK 1240
257	12	31.6	35	9	AZ867805	2M0178U13
258	12	31.6	36	11	TA203A08P	
259	12	31.6	37	1	AI446151	T. brucei
260	12	31.6	37	9	BH902529	tj07ell.x
261	12	31.6	38	9	AZ305231	SALK 0919
262	12	31.6	38	9	AZ782406	2M0022N12
263	12	31.6	38	10	AJ597895	Arabidops
264	12	31.6	39	9	AZ626090	1M0466G12
265	12	31.6	40	10	AG205568	AG205568 Oryza sat
266	12	31.6	40	8	H85898	ys92h07.gi
267	12	31.6	41	1	AV955327	AV955327
268	12	31.6	42	2	BF101499	601753311
269	12	31.6	42	6	CP643224	D61.D09.F
270	12	31.6	44	1	AV832733	AV832733
271	12	31.6	44	3	BJ001541	BJ001541
272	12	31.6	45	1	AL043028	DKFZp434P
273	12	31.6	45	10	CL640579	G077B12.G
274	12	31.6	46	1	AA206549	2G57F01.s
275	12	31.6	46	1	AA291212	2G18B12.8
276	12	31.6	46	9	AZ313975	1M0030K06
277	12	31.6	47	2	BES36289	601062691
278	12	31.6	47	10	AL760555	Arabidops
279	12	31.6	48	10	CL844507	EY12526-5
280	12	31.6	49	1	AA922626	om50a02.s
281	12	31.6	49	1	AI420261	LF08B06.x
282	12	31.6	49	2	BE960341	601653061
283	12	31.6	49	11	DMES46181	Drosophila
284	12	31.6	50	1	AUI02933	AUI02933
285	12	31.6	50	1	AUI04857	AUI04857
286	12	31.6	50	1	AUI04946	AUI04946
287	12	31.6	50	1	AUI05321	AUI05321
288	12	31.6	50	1	AUI05323	AUI05323
289	12	31.6	50	1	AUI05332	AUI05332
290	12	31.6	50	1	AUI05347	AUI05347
291	12	31.6	50	1	AUI05353	AUI05353
292	12	31.6	50	1	AUI05359	AUI05359
293	12	31.6	50	1	AUI05360	AUI05360
294	12	31.6	50	1	AUI05364	AUI05364
295	12	31.6	50	1	AUI05365	AUI05365
296	12	31.6	50	1	AUI05374	AUI05374
297	12	31.6	50	1	AUI05381	AUI05381
298	12	31.6	50	1	AUI05392	AUI05392
299	12	31.6	50	1	AUI06152	AUI06152
300	12	31.6	50	1	AUI06191	AUI06191

ALIGNMENTS

RESULT 1
 A1544460
 LOCUS
 DEFINITION
 A1544460.1 Zebrafish WashU MPMG EST Danio rerio cDNA clone
 IMAGE:3717691.3, similar to SW:CA19 HUMAN P20849 COLLAGEN ALPHA
 1(IX) CHAIN PRECURSOR. ; mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 A1544460.1 GI:4461833
 EST.

ORIGIN

Query Match 42.6%; Score 16.2; DB 1; Length 46;
 Best Local Similarity 72.4%; Fred. No. 1.1e+05;
 Matches 21; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GGGGACTTCCGTCGGGACTTCCAGGG 29
 |||||
 DB 9 GGGGATTTCCGTTCCGGGGGTTCCGGG 37
 |||||

SOURCE
 ORGANISM
 Danio rerio (zebrafish)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 REFERENCE
 AUTHORS
 1 (bases 1 to 46)
 Clark, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
 Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
 Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
 Waterston, R. and Wilson, R.
 WASHU Zebrafish EST Project 1998
 UNPUBLISHED (1998)
 CONTACT: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrafish@watson.wustl.edu
 CDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
 Matthew Clark. DNA Sequencing by: Washington University Genome
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,
 Missouri (web address: www.genomesystems.com) (email contact:
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama
 (web address: www.reagen.com) (email contact: info@reagen.com) and
 ResourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
 www.rzpd.de)
 Trace considered overall poor quality
 Possible reversed clone: similarity on wrong strand
 Seq primer: T7 ET from Amersham
 High quality sequence stop: 1
 POLYA=No.
 FEATURES
 Location/Qualifiers
 1..46
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="IMAGE:3717691"
 /sex="mixed"
 /tissue_type="26 somite embryos, adult livers, shield
 stage embryos"
 /lab_host="XLI-blue MRF"
 /clone_lib="Zebrafish WashU MPMG EST"
 /note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; 1st
 strand cDNA was primed with a Not I - oligo(dT)15 primer
 [5'pGACTAGTCTAGATCGGAGCGGCCCTTTTCTTTT3'];
 double-stranded cDNA was ligated to Sal I adaptors (BRL),
 digested with Not I and cloned into the Not I and Sal I
 sites of the pSPORT1 vector (BRL). Library was constructed
 by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
 Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
 analysis were selected following oligonucleotide
 hybridization fingerprinting of arrayed clones from
 zebrafish late somitogenesis (26 ss), adult liver or
 embryonic shield stage (5.6 h) libraries. Fingerprint
 data were used to computationally cluster cDNAs, and a
 single cDNA from each cluster was chosen for sequencing.
 In some cases multiple members of the same cluster were
 sequenced to assess clustering parameters or single clones
 were sequenced additional times to assess quality
 control."

```
RESULT 2
AA878803/c
LOCUS
DEFINITION
  AA878803      25 bp    mRNA      linear      EST 14-APR-1998
  of86e04.s1 NCI CGAP Li5 Homo sapiens cDNA clone IMAGE:1437246 3'
  similar to SW1PRPM HUMAN P10161 SALIVARY PROLINE-RICH PROTEIN PO
  ;contains TARI.t3 TARI repetitive element ;, mRNA sequence.
ACCESSION
AA878803
VERSION
AA878803.1 GI:2987768
KEYWORDS
EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Homnidae; Homo.
REFERENCE
  1 (bases 1 to 25)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs@mail.nih.gov
  Trace considered overall poor quality
  Insert Length: 185 Std Error: 0.00
  Seq primer: -40m13 fwd. RT from Amersham
  High quality sequence stop: 1.
  Location/Qualifiers
    1..25
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:1437246"
      /tissue_type="hepatic adenoma"
      /lab_host="DH10B"
      /clone_lib="NCI CGAP Li5"
      /notes="NotI; Cloned unidirectionally. Primer: Oligo dt.
      Average insert size 0.8 kb."
ORIGIN
  source
    Query Match 42.1%; Score 16; DB 1; Length 25;
    Best Local Similarity 79.2%; Pred. No. 1.3e+05;
    Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
  Qy 10 CCGCTGGGACTTTCACGGGGAC 33
  Db 25 CGGCGGGGCTTTCCGGGGAC 2
RESULT 3
AG198773/c
LOCUS
DEFINITION
  AG198773      30 bp    DNA      linear      GSS 06-MAR-2004
  Pan troglodytes DNA, clone: RP43-079M17.T7, genomic survey
  sequence.
ACCESSION
AG198773
VERSION
AG198773.1 GI:45230949
KEYWORDS
GSS.
SOURCE
  Pan troglodytes (chimpanzee)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Homnidae; Pan.
REFERENCE
  1
  Park.H., Kim.Y., Kim.S., Han.Y., Woo.T., Park.K., Eun.C.J.,
  Hoon.S.T., Chu.M., Kim.H., Joo.S., Kim.C., Song.W. and Yoo.H.
  BAC end Sequences of Library RP-43
  Unpublished
  2 (bases 1 to 30)
  Park.H., Kim.Y., Kim.S., Han.Y., Woo.T., Park.K., Eun.C.J.,
  Hoon.S.T., Chu.M., Kim.H., Joo.S., Kim.C., Song.W. and Yoo.H.
  Direct Submission
  Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of
  Bioscience and Biotechnology (KRIBB), Genome Research Center (GRC);
52, Oun-dong, Yusong-gu, Daejeon 305-333, Korea
(E-mail:redstone@mail.kribb.re.kr, URL:http://phs.grc.kribb.re.kr/,
Tel:82-42-866-7181, Fax:82-42-860-4409)
Clones are derived from the chimpanzee BAC library RP-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
  Sequencing: T7
LIBRARY
  Vector : pBACe3.6
  R.Site 1 : ECORI
  R.Site 2 : ECORI.
  Location/Qualifiers
    1..30
      /organism="Pan troglodytes"
      /mol_type="genomic DNA"
      /db_xref="taxon:9598"
      /clone="RP43-079M17.T7"
      /sex="male"
      /cell_type="lymphocytes"
      /clone_lib="RP-43 Chimpanzee Male BAC Library"
FEATURES
  source
    Query Match 42.1%; Score 16; DB 10; Length 30;
    Best Local Similarity 79.2%; Pred. No. 1.3e+05;
    Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
  Qy 6 CTTTCGGTGGGACTTTCACGGG 29
  Db 29 CTTTCCTCGAGGCGCTTTCACGGG 6
RESULT 4
CC178684/c
LOCUS
DEFINITION
  CC178684      49 bp    mRNA      linear      GSS 02-MAY-2003
  RST325 BayGenomics Gene Trap Library pGT2MpfA Mus musculus CDNA,
  mRNA sequence.
ACCESSION
CC178684
VERSION
CC178684.1 GI:30317235
KEYWORDS
GSS.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 49)
  BayGenomics.
  http://baygenomics.ucsf.edu/
  Unpublished (2001)
  Contact: BayGenomics
  Bay Area Functional Genomics Consortium (BayGenomics)
  Email: info@baygenomics.ucsf.edu
  Sequence tag generated by 5' RACE of total RNA from gene trap ES
  cell line. ES cell lines harboring insertion mutation of target
  gene are available upon request from BayGenomics. Annotation
  information available from
  http://baygenomics.ucsf.edu/cgi-bin/BaySearch.py?OPTION=EXACT&TYPE=
  CELL_LINE&KEY=RST325
  Class: Gene Trap.
  Location/Qualifiers
    1..49
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="129 ola"
      /db_xref="taxon:10090"
      /sex="Male"
      /cell_type="Embryonic stem cell"
      /clone_lib="BayGenomics Gene Trap Library pGT2MpfA"
      /notes="Vector: pGT2MpfA"
ORIGIN
  source
    Query Match 42.1%; Score 16; DB 9; Length 49;
    Best Local Similarity 79.2%; Pred. No. 1.4e+05;
```

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 GACTTTCCGCTGGGACTTTCCAG 27
 |||||
 Db 26 GACTTTCCGCTGGGCTGCACCG 3

RESULT 5
 R75772 37 bp mRNA linear EST 06-JUN-1995
 LOCUS YL21a11_r1 Soares breast 2NdbSet Homo sapiens cDNA clone
 DEFINITION IMAGE:I58876 5' similar to SP:A42445 A42445 FSP27=FAT-SPECIFIC GENE
 FSP27 - ; mRNA sequence.

ACCESSION R75772
 VERSION R75772.1 GI:850454
 KEYWORDS EST.
 ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 37)
 AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.

TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

Insert Size: 716
 High quality sequence starts: 1
 High quality sequence stops: 1
 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Trace considered overall poor quality
 Possible reversed clone: similarity on wrong strand
 Insert Length: 716 Std Error: 0.00
 Seq primer: M13RP1
 High quality sequence stop: 1.

FEATURES
 source Location/Qualifiers
 1..37
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:572930"
 /db_xref="taxon:9606"
 /clone="IMAGE:I58876"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares breast 2NdbSet"
 /note="Organ: breast; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; let strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCCCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 230. Library constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match 41.1%; Score 15.6; DB 8; Length 37;
 Best Local Similarity 81.8%; Pred. No. 1.9e+05;
 Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY
 Db

RESULT 6
 AI416657
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Glycine max (soybean)
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

REFERENCE
 AUTHORS

1 (bases 1 to 28)
 Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)

TITLE
 JOURNAL
 COMMENT

Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

When it has been determined, an EST from the other end of this clone is listed in the 'Other ESTs on clone' field. Other ESTs: sal0b04.y1 GENOME SYSTEMS CLONE ID: Gm-cl003-8 Trace considered overall poor quality Possible reversed clone: similarity on wrong strand This clone is available through: Biogenetic Services, 801 32nd Ave. Brookings, SD 57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
 Seq primer: -40UP from Gibco
 High quality sequence stop: 1

POLYA-No. Location/Qualifiers
 1..28
 /organism="Glycine max"
 /mol_type="mRNA"
 /cultivar="Williams"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl003-8"
 /tissue_type="immature cotyledon"
 /lab_host="XL10-Gold"
 /clone_lib="Gm-cl003"
 /note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2: XhoI; Immature Cotyledon cDNA. The mRNA was isolated and then Stratagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA is hemimethylated. Stratagene's first-strand synthesis primer was used (GAGAGAGAGAGAGAGAACTAGTCTCGAG(T)-18). After second-strand synthesis, the cDNA ends were 'polished' with clone Pfu DNA polymerase, ligated to EcoRI adapters, and phosphorylated. The XhoI site within the first-strand synthesis primer was restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using GibcoBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into Stratagene's pBluescript II XR Predigested vector (pBluescript II SK(+)) that had been

FEATURES
 source

chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 40.5%; Score 15.4; DB 9; Length 34;
 Best Local Similarity 76.0%; Pred. No. 2.2e+05;
 Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 10 CCGCTGGGGACTTTCAGGGGACT 34
 ||||| ||||| ||||| ||||| |||||
 Db 1 CCCTGGGACACTTGTAGGTGACT 25

RESULT 9
 CG800335/c

LOCUS
 DEFINITION 1118010B12.2EL_xl 1118 - RescueMu Grid S Zea mays genomic, genomic survey sequence.

ACCESSION
 VERSION CG800335.1 GI:38236121

KEYWORDS
 GSS.

SOURCE
 Zea mays

ORGANISM
 Zea mays

REFERENCE 1 1118010B12.2EL_xl 1118 - RescueMu Grid S Zea mays genomic, genomic survey sequence.

AUTHORS
 Walbot V.

TITLE
 Maize genomic sequences found using engineered RescueMu transposon

JOURNAL
 COMMENT

Unpublished (2001)
 Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu

Possible ligation site of ends cut by 2 different endonucleases.
 Reverse complemented post-ligation sequence from source sequence.
 Plate: 1118010 row: 26
 Class: transposon-tagged.

FEATURES
 source

1..50
 Location/Qualifiers
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultiVar="mixed background W23/A188/B73"
 /db_xref="taxon:4577"
 /tissue_type="leaf"
 /dev_stages="adult"
 /lab_host="DH10B"
 /clone_lib="1118 - RescueMu Grid S"

/note="Organ: leaf; Vector: RescueMu (engineered from pBlueScript backbone); Site_1: BamHI; Site_2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'RescueMu.' Grid S was grown at San Diego in 2002. DNA was extracted from leaf strips, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

ORIGIN

Query Match 40.5%; Score 15.4; DB 10; Length 50;
 Best Local Similarity 66.7%; Pred. No. 2.3e+05;
 Matches 22; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 4 GACTTTCGGCTGGGACTTTCAGGGGACTTT 36
 ||||| ||||| ||||| ||||| |||||
 Db 45 GTCCTTCGGCTGGGACTTTCAGGGGACTTT 13

RESULT 10
 CG776618

LOCUS
 DEFINITION 1123002C09.2EL_xl 1123 - RescueMu Grid L Zea mays genomic, genomic survey sequence.

ACCESSION
 VERSION CG776618
 CG776618.1 GI:38034884

KEYWORDS
 GSS.

SOURCE
 Zea mays

ORGANISM
 Zea mays

REFERENCE 1 1123002C09.2EL_xl 1123 - RescueMu Grid L Zea mays genomic, genomic survey sequence.

AUTHORS
 Walbot V.

TITLE
 Maize genomic sequences found using engineered RescueMu transposon

JOURNAL
 COMMENT

Unpublished (2001)
 Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu

Possible ligation site of ends cut by 2 different endonucleases.
 Reverse complemented post-ligation sequence from source sequence.
 Plate: 1123002 row: 2
 Class: transposon-tagged.

FEATURES
 source

1..42
 Location/Qualifiers
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultiVar="mixed background W23/A188/B73/K55"
 /db_xref="taxon:4577"
 /tissue_type="leaf"
 /dev_stages="adult"
 /lab_host="DH10B"

/clone_lib="1123 - RescueMu Grid L"
 /note="Organ: leaf; Vector: RescueMu (engineered from pBlueScript backbone); Site_1: BamHI; Site_2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'RescueMu.' Grid L was grown in Molokai in 2001. DNA was extracted from leaf strips, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

ORIGIN

Query Match 39.5%; Score 15; DB 10; Length 42;
 Best Local Similarity 67.7%; Pred. No. 3.2e+05;
 Matches 21; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 GGGGACTTTCGGCTGGGACTTTCAGGGGG 31
 ||||| ||||| ||||| ||||| |||||
 Db 5 GTGGATCTTCCCTGGGACTTTCAGGGGG 35

RESULT 11
 AI917121/c

LOCUS
 DEFINITION t952a12.x1 NCI CGAP Kid8 Homo sapiens cDNA clone IMAGE:232190 3', similar to SW:CA13_BOVIN P04258 COLLAGEN ALPHA 1(III) CHAIN.

ACCESSION
 VERSION AI917121.1 GI:5636976

KEYWORDS
 EST.

SOURCE
 Homo sapiens (human)

ORGANISM
 Homo sapiens


```

RESULT 19
CZ442924/c
LOCUS IB33G02.fwd HIV-vector integration sites from well-expressed
DEFINITION proviruses in human Jurkat T cells Homo sapiens genomic clone
IB33G02.fwd, genomic survey sequence.
ACCESSION CZ442924
VERSION 1
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 35)
AUTHORS Lewinski,M.K., Blsgrove,D., Shinn,P., Chen,H., Hoffmann,C.,
Hannenhalli,S., Verdin,E., Berry,C.C., Ecker,J.R. and Bushman,F.D.
TITLE Genome-wide analysis of chromosomal features repressing human
JOURNAL immunodeficiency virus transcription
PUBMED J. Virol. 79 (11), 6610-6619 (2005)
COMMENT 15890899
Contact: Bushman FD
Department of Microbiology
University of Pennsylvania School of Medicine
402C Johnson Pavilion, 3610 Hamilton Walk, Philadelphia, PA
19104-6076 USA
Tel: 215 573 8732
Fax: 215 573 4856
Email: bushman@mail.med.upenn.edu
Class: PCR with specific primers.
FEATURES
source
1..35
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="IB33G02.fwd"
/cell_line="Jurkat"
/clone_lib="HIV-vector integration sites from
well-expressed proviruses in human Jurkat T cells"
/note="Vector: LTR-Tat-IRES-GFP (pEV731); We have
investigated regulatory sequences in noncoding human DNA
that are associated with repression of an integrated human
immunodeficiency virus type 1 (HIV-1) promoter. HIV-1
integration results in the formation of precise and
homogeneous junctions between viral and host DNA, but
integration takes place at many locations. Thus, the
variation in HIV-1 gene expression at different
integration sites reports the activity of regulatory
sequences at nearby chromosomal positions. Negative
regulation of HIV transcription is of particular interest
because of its association with maintaining HIV in a
latent state in cells from infected patients. To identify
chromosomal regulators of HIV transcription, we infected
Jurkat T cells with an HIV-based vector transducing green
fluorescent protein (GFP) and separated cells into
populations containing well-expressed (GFP-positive) or
poorly expressed (GFP-negative) proviruses. We then
determined the chromosomal locations of the two classes by
sequencing 971 junctions between viral and cellular DNA.
Possible effects of endogenous cellular transcription were
characterized by transcriptional profiling. Low-level GFP
expression correlated with integration in (i) gene
deserts, (ii) centromeric heterochromatin, and (iii) very
highly expressed cellular genes. These data provide a
genome-wide picture of chromosomal features that repress
transcription and suggest models for transcriptional
latency in cells from HIV-infected patients."
ORIGIN
Query Match 37.4%; Score 14.2; DB 10; Length 35;
Best Local Similarity 70.4%; Pred. No. 6.4e+05;
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

CZ442924
IB33G02.fwd HIV-vector integration sites from well-expressed
DEFINITION proviruses in human Jurkat T cells Homo sapiens genomic clone
IB33G02.fwd, genomic survey sequence.
ACCESSION CZ442924
VERSION 1
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 35)
AUTHORS Lewinski,M.K., Blsgrove,D., Shinn,P., Chen,H., Hoffmann,C.,
Hannenhalli,S., Verdin,E., Berry,C.C., Ecker,J.R. and Bushman,F.D.
TITLE Genome-wide analysis of chromosomal features repressing human
JOURNAL immunodeficiency virus transcription
PUBMED J. Virol. 79 (11), 6610-6619 (2005)
COMMENT 15890899
Contact: Bushman FD
Department of Microbiology
University of Pennsylvania School of Medicine
402C Johnson Pavilion, 3610 Hamilton Walk, Philadelphia, PA
19104-6076 USA
Tel: 215 573 8732
Fax: 215 573 4856
Email: bushman@mail.med.upenn.edu
Class: PCR with specific primers.
FEATURES
source
1..35
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="IB33G02.fwd"
/cell_line="Jurkat"
/clone_lib="HIV-vector integration sites from
well-expressed proviruses in human Jurkat T cells"
/note="Vector: LTR-Tat-IRES-GFP (pEV731); We have
investigated regulatory sequences in noncoding human DNA
that are associated with repression of an integrated human
immunodeficiency virus type 1 (HIV-1) promoter. HIV-1
integration results in the formation of precise and
homogeneous junctions between viral and host DNA, but
integration takes place at many locations. Thus, the
variation in HIV-1 gene expression at different
integration sites reports the activity of regulatory
sequences at nearby chromosomal positions. Negative
regulation of HIV transcription is of particular interest
because of its association with maintaining HIV in a
latent state in cells from infected patients. To identify
chromosomal regulators of HIV transcription, we infected
Jurkat T cells with an HIV-based vector transducing green
fluorescent protein (GFP) and separated cells into
populations containing well-expressed (GFP-positive) or
poorly expressed (GFP-negative) proviruses. We then
determined the chromosomal locations of the two classes by
sequencing 971 junctions between viral and cellular DNA.
Possible effects of endogenous cellular transcription were
characterized by transcriptional profiling. Low-level GFP
expression correlated with integration in (i) gene
deserts, (ii) centromeric heterochromatin, and (iii) very
highly expressed cellular genes. These data provide a
genome-wide picture of chromosomal features that repress
transcription and suggest models for transcriptional
latency in cells from HIV-infected patients."
ORIGIN
Query Match 37.4%; Score 14.2; DB 10; Length 35;
Best Local Similarity 70.4%; Pred. No. 6.4e+05;
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 12 GCTGGGACTTTCAGGGGACTTTCC 38
DB 27 GGTGAGTACTCTGCAGGAGGACGTTAC 1

RESULT 20
TA379B06Q/c
LOCUS TA379B06Q 37 bp DNA linear GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 379B06, reverse sequence,
genomic survey sequence.
ACCESSION AL497861
VERSION AL497861.1 GI:11873583
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 37)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.E., Rajandream,M.A. and Barrell,B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES
source
1..37
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="379B06"
ORIGIN
Query Match 37.4%; Score 14.2; DB 11; Length 37;
Best Local Similarity 70.4%; Pred. No. 6.4e+05;
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 12 GCTGGGACTTTCAGGGGACTTTCC 38
DB 36 GGTGGGATTTTCATGTTGATTTCC 10

RESULT 21
AI039253/c
LOCUS AI039253 40 bp mRNA linear EST 28-AUG-1998
DEFINITION ox33b01.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone
IMAGE:1658089 3', similar to SW.PF11_FIG P51524 PROPHENIN-1
PRECURSOR :, mRNA sequence.
ACCESSION AI039253
VERSION AI039253.1 GI:3278447
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 40)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

```

```

TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs@mail.nih.gov
           This clone is available royalty-free through LNL; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           Trace considered overall poor quality
           Insert length: 1067 Std Error: 0.00
           Seq primer: -40ml3 fwd. RT from Amersham
           High quality sequence stop: 1.
FEATURES   Location/Qualifiers
            1..40
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:1658089"
            /dev_stage="8-9 weeks"
            /lab_host="DH10B"
            /clone_lib="Soares total fetus Nb2HF8_9w"
            /note="Vector: p7T3D-Pac (Pharmacia) with a modified
            polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
            was prepared from mRNA obtained from pooled 8-9 week
            (total) fetus material with a Not I - oligo(dT) primer [5'
            TGTTACCATCTGAAGTGGAGCGCGCTTAATTTTCTTTT 3'].
            Double-stranded cDNA was ligated to Eco RI adaptors
            (Pharmacia), digested with Not I and cloned into the
            Not I and Eco RI sites of the modified p7T3 vector. Library
            went through one round of normalization, and was
            constructed by Bento Soares and M. Fatima Ronaldo."
ORIGIN
Query Match      37.4%; Score 14.2; DB 1; Length 40;
Best Local Similarity 70.4%; Pred. No. 6.5e+05;
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY      3  GGACTTTCGCTGGGACTTTCAGGG 29
      ||||| ||||| ||||| ||||| |||||
Db      28  GGTTTCCGGGGGGGCCCTTCTAGG 2

RESULT 22
AUI04208/c
LOCUS      AUI04208      50 bp      mRNA      linear      EST 28-JAN-2004
DEFINITION AUI04208 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            ADSE00350, mRNA sequence.
ACCESSION  AUI04208
VERSION     AUI04208.1 GI:13553729
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE  1 (bases 1 to 50)
AUTHORS   Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J.,
            Hata,H., Oka,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K.,
            Sakaki,Y., Nakamura,Y., Suyama,A. and Sugano,S.
            Diverse transcriptional initiation revealed by fine, large-scale
            mapping of mRNA start sites
JOURNAL    EMBO Rep. 2 (5), 388-393 (2001)
PUBMED     11375929
COMMENT    Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: yusuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and
            Sugano,S. Construction and characterization of a full
            length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2),
            149-156 (1997).
FEATURES   Location/Qualifiers
            1..50

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="ADSE00350"
/clone_lib="Sugano Homo sapiens cDNA library"

ORIGIN
Query Match      37.4%; Score 14.2; DB 1; Length 50;
Best Local Similarity 70.4%; Pred. No. 6.7e+05;
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY      1  GGGGACTTTCGCTGGGACTTTCAG 27
      ||||| ||||| ||||| ||||| |||||
Db      30  GGGGATTGGCGCTGGGCTCTTCCCTG 4

RESULT 23
CC459460
LOCUS      CC459460      50 bp      DNA      linear      GSS 30-MAY-2003
DEFINITION SALK_129954.28.55.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_129954.28.55.x, genomic
            survey sequence.
ACCESSION  CC459460
VERSION     CC459460.1 GI:31225746
KEYWORDS   GSS
SOURCE     Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicaceae; Arabidopsi.
REFERENCE  1 (bases 1 to 50)
AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
            Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
            Shinn,P., Zimmerman,J. and Ecker,J.R.
            A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
            Unpublished (2001)
            Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGNAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
            Location/Qualifiers
            1..50
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /ecotype="Col-0"
            /db_xref="taxon:3702"
            /clone="SALK_129954.28.55.x"
            /clone_lib="Arabidopsis thaliana TDNA insertion lines"
            /note="PCR was performed on Arabidopsis thaliana lines
            each of which contains one or more TDNA insertion
            elements. the resultant fragment for each line was
            directly sequenced to determine the genomic sequence at
            the site of insertion. Details of the protocols used can
            be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match      37.4%; Score 14.2; DB 9; Length 50;
Best Local Similarity 62.9%; Pred. No. 6.7e+05;
Matches 22; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY      1  GGGGACTTTCGCTGGGACTTTCAGGGGACTT 35
      ||||| ||||| ||||| ||||| |||||
Db      12  GGGGAAGTTGGTTGGCGATTTCGATGAAATT 46

RESULT 24

```


COMMENT

Clones are derived from the chimpanzee BAC library RP-43 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS

Sequencing: TJ

LIBRARY

Vector : pBACe3.6

R-site 1 : EcorI

R-site 2 : EcorI

Location/Qualifiers

1..38

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

/clone="RP43-065002.TJ"

/sex="male"

/cell_type="lymphocytes"

/clone_lib="RP-43 Chimpanzee Male BAC Library"

FEATURES

source

ORIGIN

Query Match 36.8%; Score 14; DB 10; Length 38;
Best Local Similarity 66.7%; Pred. No. 7.7e+05;
Matches 20; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 3 GGACTTTCGCTGGGACATTTCCAGGGGGA 32

||||| | | | | | | | | | | | | | | |

Db 8 GGNATATTGGGGGAATATTCTAGGGGGA 37

RESULT 27

AZ595836

LOCUS

DEFINITION 1M0408008R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0408008 R, genomic survey sequence.

ACCESSION AZ595836

VERSION 1 GI:11718026

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 40)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)

JOURNAL

COMMENT

Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0408 row: 0 column: 08

Seq primer: CACACAGAAACAGTATGACC

Class: plasmid ends

High quality sequence stop: 40.

Location/Qualifiers

1..40

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0408008"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, Tl-resistant, P-"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

FEATURES

source

1..40

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0545P18"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, Tl-resistant, P-"

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 36.8%; Score 14; DB 9; Length 40;
Best Local Similarity 66.7%; Pred. No. 7.8e+05;
Matches 20; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GGGGACTTTCGCTGGGACATTTCCAGGGG 30

||||| | | | | | | | | | | | | | | |

Db 11 GGGGCTGTGCAGCTTGTTCCTTCAGTGG 40

RESULT 28

AZ664762

LOCUS

DEFINITION AZ664762 40 bp DNA linear GSS 14-DEC-2000 1M0545P18F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0545P18 F, genomic survey sequence.

ACCESSION AZ664762

VERSION

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 40)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)

JOURNAL

COMMENT

Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0545 row: P column: 18

Seq primer: CGTTGTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 40.

Location/Qualifiers

1..40

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0545P18"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, Tl-resistant, P-"

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert length: 672 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.

FEATURES

source

```
1..49
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2011857"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/clone_lib="NCI CGAP Brn23"
/note="Organ: brain; Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGGAGCGCGCATATCTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaudo."
```

ORIGIN

Query Match 36.8%; Score 14; DB 1; Length 49;
Best Local Similarity 60.5%; Pred. No. 8e+05;
Matches 23; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
QY 1 GGGGACTTCCGCTGGGACTTTCACAGGGGACTTTC 38
|||||
DB 39 GGGGTTTTCCCTTTCCCTTTTGGGGGTTTTTCC 2
```

RESULT 34

```
BE311372
LOCUS BE311372
DEFINITION BE311372
ACCESSION BE311372.1 GI:9175271
VERSION BE311372.1
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM7650 row: e column: 19

High quality sequence start: 3

High quality sequence stop: 50.

Location/Qualifiers

1..50

/organism="Mus musculus"

/mol_type="mRNA"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:3158010"

FEATURES

source

Query Match 36.3%; Score 13.8; DB 1; Length 28;
Best Local Similarity 72.0%; Pred. No. 8.8e+05;
Matches 18; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GGGGACTTTCGCTGGGACTTTC 25

```
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"
```

ORIGIN

Query Match 36.8%; Score 14; DB 2; Length 50;
Best Local Similarity 77.3%; Pred. No. 8e+05;
Matches 17; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
QY 8 TTCCGCTGGGACTTTCACGGG 29
|||||
DB 4 TACCGCTCCGAGTTTCCCGG 25
```

RESULT 35

```
AI432930
LOCUS AI432930
DEFINITION ch4c11.x1 NCI CGAP Lym12 Homo sapiens cDNA clone IMAGE:2121140 3' similar to SW:PRP2 HUMAN P02812 SALIVARY PROLINE-RICH PROTEIN PRECURSOR ; contains element MSR1 repetitive element ;, mRNA sequence.
ACCESSION AI432930
VERSION AI432930.1 GI:4285621
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 28)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Life Technologies catalog #: 11547-015
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

REFERENCE

1 (bases 1 to 28)
AUTHORS NCI-CGAP
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Life Technologies catalog #: 11547-015
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

FEATURES

source

```
1..28
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2121140"
/tissue_type="lymphoma, follicular mixed small and large cell"
/lab_host="DH10B"
/clone_lib="NCI CGAP Lym12"
/note="Organ: lymph node; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.25 kb. Life Technologies catalog #: 11547-015"
```

ORIGIN

Query Match 36.3%; Score 13.8; DB 1; Length 28;
Best Local Similarity 72.0%; Pred. No. 8.8e+05;
Matches 18; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GGGGACTTTCGCTGGGACTTTC 25

Db 1 GGGGAATTTCCCGGGGGCCCCC 25

||||| ||||| ||||| ||||| |||||

RESULT 36
AZ628058 33 bp DNA linear GSS 13-DEC-2000
LOCUS 1M0476122F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0476122 F, genomic survey sequence.
ACCESSION AZ628058
VERSION AZ628058.1 GI:11750248
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 33)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: rdunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0476 row: 1 column: 22
Seq primer: CGTTGTAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 33.

FEATURES
Location/Qualifiers
1..33
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0476122"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWB42 [gi|4732114|gb|AF129072.1], a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN
Query Match 36.3%; Score 13.8; DB 9; Length 33;
Best Local Similarity 72.0%; Pred. No. 9e+05; Length 33;
Matches 18; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 9 TCCGCTGGGGACTTTCCAGGGGGAC 33
||||| ||||| ||||| ||||| |||||

Db 9 TCCCATGGGGAGTAAAGGGGGAC 33
||||| ||||| ||||| ||||| |||||

RESULT 37
AI579983/c 37 bp mRNA linear EST 14-DEC-1999
LOCUS tq45a05.x1 NCI CGAP Ut1 Homo sapiens cDNA clone IMAGE:2211728 3'
DEFINITION similar to TR:Q0805 Q0805 SALIVARY PROLINE-RICH PROTEIN L
;contains element MSRI repetitive element ;, mRNA sequence.
ACCESSION AI579983
VERSION AI579983.1 GI:4564359
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 37)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 2882 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1
POLYA=No.

FEATURES
Location/Qualifiers
1..37
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2211728"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/clone_lib="NCI CGAP Ut1"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 Kb. Life Technologies catalog #:
11538-014"

ORIGIN
Query Match 36.3%; Score 13.8; DB 1; Length 37;
Best Local Similarity 72.0%; Pred. No. 9.2e+05; Length 37;
Matches 18; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 7 TTTCCGCTGGGGACTTTCCAGGGGG 31
||||| ||||| ||||| ||||| |||||

Db 36 TTTCCGGGGGGGTTTTTCGGGGGG 12
||||| ||||| ||||| ||||| |||||

RESULT 38
DR48J2S/c 41 bp DNA linear GSS 22-NOV-2002
LOCUS Danio rerio genomic clone DKEY-48J2, genomic survey sequence.
DEFINITION AL982948
VERSION AL982948.1 GI:25186174
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)

VERSION AI544460.1 GI:4461833
 KEYWORDS EST.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Neopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 46)
 AUTHORS Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, P., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 TITLE WashU Zebrafish EST Project 1998
 JOURNAL Unpublished (1998)
 COMMENT Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrafish@wustl.edu
 cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center. Clone distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address: www.rzpd.de)

Trace considered overall poor quality
 Possible reversed clone: similarity on wrong strand
 Seq primer: T7 Et from Amersham
 High quality sequence stop: 1
 POLYA=No.

FEATURES

source Location/Qualifiers
 1..46
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="IMAGE:3717691"
 /sex="mixed"
 /tissue types="26 somite embryos, adult livers, shield stage embryos"
 /lab_host="X11-blue MRP"
 /clone_lib="Zebrafish WashU MPIMG EST"
 /notes="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; 1st strand cDNA was primed with a Not I - oligo(dT)15 primer [5'pGACTAGTTCTAGATCGGAGCGCGCCCTTTTCTTTTCTTTT3']; double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

ORIGIN

Query Match 36.3%; Score 13.8; DB 1; Length 46;
 Best Local Similarity 63.6%; Pred. No. 9.5e+05;
 Matches 21; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
 Qy 1 GGGGACTTTCCTGGGGACTTTCACGGGGAC 33
 Db 34 GGGAAACCCCGAACCGGGAATCCCGGGGCC 2

RESULT 42

BJ057570
 LOCUS
 DEFINITION

BJ057570 NIBB Mochii normalized Xenopus tailbud library Xenopus laevis cDNA clone XL104h07 5', mRNA sequence.
 46 bp mRNA linear EST 29-SEP-2003

ACCESSION BJ057570.1 GI:17479651
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Xenopus laevis (African clawed frog)
 Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

1 (bases 1 to 46)
 AUTHORS Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and Kohara, Y.

Expressed genes in X. laevis embryo

Unpublished (2001)

Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp

The information of this clone is available through the following URL.

http://xenopus.nibb.ac.jp.

Location/Qualifiers

1..46

/organism="Xenopus laevis"

/mol_type="mRNA"

/db_xref="taxon:8355"

/clone="XL104h07"

/tissue_type="whole embryo"

/clon_stage="stage 25"

/clone_lib="NIBB Mochii normalized Xenopus tailbud library"

ORIGIN

Query Match 36.3%; Score 13.8; DB 3; Length 46;
 Best Local Similarity 66.7%; Pred. No. 9.5e+05;
 Matches 18; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 12 GCTGGGACTTTCACGGGGACTTTC 38

Db 14 GCTGGGAGTTGACAGGGGTTTANCC 40

RESULT 43

AU102496

LOCUS

DEFINITION

AU102496 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone HEP22967, mRNA sequence.

ACCESSION AU102496

VERSION AU102496.1 GI:13552016

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

REFERENCE 1 (bases 1 to 50)

AUTHORS Suzuki, Y., Taira, H., Tsunoda, T., Mizushima-Sugano, J., Sese, J.,

Hata, H., Ota, T., Isogai, T., Tanaka, T., Morishita, S., Okubo, K.,

Sakaki, Y., Nakamura, Y., Suyama, A., and Sugano, S.

Diverse transcriptional initiation revealed by fine, large-scale

mapping of mRNA start sites

EMBO Rep. 2 (5), 388-393 (2001)

11375929

COMMENT Contact: Yutaka Suzuki

Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: ysuzuki@ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano,S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES

source
1..50
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEP22967"
/clone_lib="Sugano Homo sapiens cDNA library"

ORIGIN

Query Match 36.3%; Score 13.8; DB 1; Length 50;
Best Local Similarity 63.6%; Pred. No. 9.6e+05;
Matches 21; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 6 CTTTCGGCTGGGACTTTCAGGGGACTTCC 38
|||||
Db 2 CTTTCGGCTGGGCTGTTTCTCGCAGGTTTC 34
|||||

RESULT 44

AU107397/c
LOCUS AU107397 Sugano Homo sapiens cDNA library EST 28-JAN-2004
DEFINITION LNC13032, mRNA sequence.
ACCESSION AU107397
VERSION AU107397.1 GI:13556918
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 50)
Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata,H., Oca,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki,Y., Nakamura,Y., Suyama,A. and Sugano,S.
Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites
JOURNAL EMBO Rep. 2 (5), 388-393 (2001)
PUBMED 11375929
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: ysuzuki@ims.u-tokyo.ac.jp

REFERENCE

AUTHORS Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata,H., Oca,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki,Y., Nakamura,Y., Suyama,A. and Sugano,S.
TITLE Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites
JOURNAL EMBO Rep. 2 (5), 388-393 (2001)
PUBMED 11375929
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: ysuzuki@ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano,S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES

source
1..50
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="LNC13032"
/clone_lib="Sugano Homo sapiens cDNA library"

ORIGIN

Query Match 36.3%; Score 13.8; DB 1; Length 50;
Best Local Similarity 63.6%; Pred. No. 9.6e+05;
Matches 21; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 GGGGACTTTCGGCTGGGACTTTCAGGGGAC 33
|||||
Db 49 GGAGACAATTCCTTGTCTGACTGTCAGGCAGTC 17
|||||

RESULT 45

CR178464/c
LOCUS CR178464 50 bp DNA linear GSS 06-JUL-2004
DEFINITION Forward strand read from insert in 3'HPRT insertion targeting and chromosome engineering clone MHP340e12, genomic survey sequence.
ACCESSION CR178464
VERSION CR178464.1 GI:49957313
KEYWORDS GSS; genome survey sequence; MICER.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.
1 (bases 1 to 50)
Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L., Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y., Rogers,J. and Bradley,A.
Direct Submission
Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. http://www.sanger.ac.uk/MICER

TITLE

JOURNAL

FEATURES

source
1..50
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHP340e12"
/clone_lib="MHPp"

ORIGIN

Query Match 36.3%; Score 13.8; DB 1; Length 50;
Best Local Similarity 88.2%; Pred. No. 9.6e+05;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 12 GCTGGGACTTCCAGG 28
|||||
Db 38 GCTGGGACTTCCAGG 22
|||||

RESULT 46

BH812215
LOCUS BH812215 32 bp DNA linear GSS 02-MAY-2002
DEFINITION SALUK_061414 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_061414, genomic survey sequence.
ACCESSION BH812215
VERSION BH812215.1 GI:20390670
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 32)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: eckersalk.edu
This is single pass sequence recovered from the left border of TDNA.

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..32
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"

/ecotype="Col-0"
 /db_xref="taxon:3702"
 /clone="SALK_06414"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /notes=PCR was performed on Arabidopsis thaliana lines
 each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html

ORIGIN

Query Match 35.8%; Score 13.6; DB 9; Length 32;
 Best Local Similarity 80.0%; Pred. No. 1.1e+06;
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 ACTTTCCGCTGGGAGCTTTC 24
 ||||| |||||
 Db 9 ACTTTCCGTAGGATCTTTC 28

RESULT 47

AZ775335 40 bp DNA linear GSS 16-FEB-2001
 LOCUS 2M0007G07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC2M0007G07 R, genomic survey sequence.

ACCESSION AZ775335
 VERSION
 KEYWORDS
 SOURCE GSS.
 ORGANISM Mus musculus (house mouse)

REFERENCE
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
 Reilly,M., Rose,R., Stokes,R., Tingley,A., von
 Niederhausern,A. and Wright,D. Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0007 row: G column: 07
 Seq primer: CACACAGGAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 40.

FEATURES

source
 1..40
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0007G07"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /note="Vector: FWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN

Query Match 35.8%; Score 13.6; DB 9; Length 40;
 Best Local Similarity 67.9%; Pred. No. 1.1e+06;
 Matches 19; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 GGGGACCTTTCCGCTGGGAGCTTTCCAGG 28
 ||||| ||||| ||||| |||||
 Db 7 GGGGGCTTGCAGTTGGGTTCTGCGCTGG 34

RESULT 48

TA26H11Q 41 bp DNA linear GSS 13-DEC-2000
 LOCUS T. brucei sheared genomic DNA clone 26h11, reverse sequence,
 genomic survey sequence.

ACCESSION AL452967
 VERSION AL452967.1 GI:11850952
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei
 ORGANISM Trypanosoma brucei
 Eukaryota; Euklenozoa; Kinetoplastida, Trypanosomatidae;
 Trypanosoma.

REFERENCE
 AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
 Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
 Melville,S.E., Rajandream,M.A. and Barrell,B.G.
 Direct Submission
 Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
 project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
 Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
 nh1@sanger.ac.uk
 Constructed at the Institute for Genomic Research (TIGR),
 Rockville, MD. Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
 to give a tight size distribution (
 4 kb). The v + i method used for the library construction is
 described in detail in Smith, H. and Venter, J.C. (Making small
 insert libraries for whole genome shotgun sequencing projects. In
 Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999).
 Email: nelsayed@tigr.org
 Details of T. brucei sequencing at the Sanger Centre are available
 at http://www.sanger.ac.uk/projects/T_brucei/.

FEATURES

source
 1..41
 /organism="Trypanosoma brucei"
 /mol_type="genomic DNA"
 /strain="TREU927"
 /db_xref="taxon:5691"
 /clone="26h11"

ORIGIN

Query Match 35.8%; Score 13.6; DB 11; Length 41;
 Best Local Similarity 67.9%; Pred. No. 1.1e+06;
 Matches 19; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 GACTTTCCGCTGGGAGCTTTCCAGGGGG 31
 ||||| ||||| ||||| |||||
 Db 13 GAGTGTCTGCAGGCGCATTTGCCAGGGGG 40

```
RESULT 49
W10989/c
LOCUS
DEFINITION
ma41b12.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:313247 5', mRNA sequence.
ACCESSION
W10989
VERSION
W10989.1 GI:1285294
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 45)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
TITLE
Unpublished (1996)
JOURNAL
Contact: Marra M/Mouse EST Project
COMMENT
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI: 203863
Seq primer: -21M13
High quality sequence stop: 1.
FEATURES
Location/Qualifiers
1..45
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:313247"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares mouse p3NMF19.5"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCATCTGAGTGGAGCGCCGCAATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
```

```
Query Match 35.8%; Score 13.6; DB 8; Length 45;
Best Local Similarity 67.9%; Pred. No. 1.1e+06;
Matches 19; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 8 TTCCGCTGGGACTTTCACGGGGGACTT 35
||| ||| ||| ||| ||| ||| |||
Db 28 TTGAGCTTGGAACTGTCCACTGGAACTT 1

RESULT 50
A1635210/c
LOCUS
DEFINITION
t22e06.x1 NCI CGAP ut2 Homo sapiens cDNA clone IMAGE:2289346 3'
similar to SW:EXTN_DAUCA P06599 EXTENSIN PRECURSOR. ;, mRNA
sequence.
ACCESSION
A1635210
VERSION
A1635210.1 GI:4686540
KEYWORDS
EST.
```

```
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 46)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
TITLE Unpublished (1997)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Trace considered overall poor quality
Insert Length: 1695 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
FEATURES
Location/Qualifiers
1..46
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2289346"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/clone_lib="NCI CGAP Ut2"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"
ORIGIN
Query Match 35.8%; Score 13.6; DB 1; Length 46;
Best Local Similarity 61.1%; Pred. No. 1.1e+06;
Matches 22; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1 GGGGACTTTCGCTGGGACTTTCACGGGGGACTTT 36
||| ||| ||| ||| ||| ||| |||
Db 41 GGGGTTTTTTCGGTGGGGGAAACCTTGGAAACCTT 6

Search completed: February 15, 2006, 21:10:37
Job time : 2541.47 secs
```

THIS PAGE BLANK (USPTO)